MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 8: 489–518 (2024)
DOI: 10.3897/mbmg.8.128689

Forum Paper

The MIEM guidelines: Minimum information for reporting of environmental metabarcoding data

Katy E. Klymus[1], Jacoby D. Baker[2], Cathryn L. Abbott[3], Rachel J. Brown[4], Joseph M. Craine[5], Zachary Gold[6], Margaret E. Hunter[7], Mark D. Johnson[8,9], Devin N. Jones[10], Michelle J. Jungbluth[11], Sean P. Jungbluth[11], Yer Lor[12], Aaron P. Maloy[13], Christopher M. Merkes[12], Rachel Noble[14], Nastassia V. Patin[15,16], Adam J. Sepulveda[10], Stephen F. Spear[12], Joshua A. Steele[16], Miwa Takahashi[17], Alison W. Watts[18], Susanna Theroux[16]

1 *U.S. Geological Survey, Columbia Environmental Research Center, Columbia, Missouri, USA*
2 *Monterey Bay Aquarium Research Institute, Moss Landing, California, USA*
3 *Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, V9T 6N7, Canada*
4 *U.S. Fish and Wildlife, Whitney Genetics Laboratory, Onalaska, Wisconsin, USA*
5 *Jonah Ventures, Boulder, Colorado, USA*
6 *NOAA Pacific Marine Environmental Laboratory, Seattle, Washington, USA*
7 *U.S. Geological Survey, Wetland and Aquatic Research Center, Gainesville, Florida, USA*
8 *Engineer Research and Development Center, Champaign, Illinois, USA*
9 *Illinois Natural History Survey, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA*
10 *U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, Montana, USA*
11 *San Francisco State University, Estuary & Ocean Science Center, Tiburon, California, USA*
12 *U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin, USA*
13 *U.S. Fish and Wildlife Service, Northeast Fishery Center, Lamar, Pennsylvania, USA*
14 *Departments of Earth, Marine, and Environmental Sciences and Environmental Sciences and Engineering, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA*
15 *Integrated Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA*
16 *Southern California Coastal Water Research Project, Costa Mesa, California, USA*
17 *Environomics Future Science Platform, Commonwealth Scientific and Industrial Research Organisation, Indian Oceans Marine Research Centre, Crawley, Western Australia, Australia*
18 *Department of Civil and Environmental Engineering, University of New Hampshire, Durham, New Hampshire, USA*

Corresponding author: Katy E. Klymus (kklymus@usgs.gov)

Abstract

Environmental DNA (eDNA) and RNA (eRNA) metabarcoding has become a popular tool for assessing biodiversity from environmental samples, but inconsistent documentation of methods, data and metadata makes results difficult to reproduce and synthesise. A working group of scientists have collaborated to produce a set of minimum reporting guidelines for the constituent steps of metabarcoding workflows, from the physical layout of laboratories through to data archiving. We emphasise how reporting the suite of data and metadata should adhere to findable, accessible, interoperable and reproducible (FAIR) data standards, thereby providing context for evaluating and understanding study results. An overview of the documentation considerations for each workflow step is presented and then summarised in a checklist that can accompany a published study or report. Ensuring workflows are transparent and documented is critical to reproducible research and should allow for more efficient uptake of metabarcoding data into management decision-making.

Key words: eDNA, eRNA, metadata, quality control, FAIR data standards, reproducibility

This article is part of:
Towards standardized molecular biodiversity monitoring
Edited by Teresita Porter, John Darling, Kelly Goodwin, Tiina Laamanen, Kristian Meissner, Toshifumi Minamoto

OPEN ACCESS

Copyright: This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

Academic editor:
Toshifumi Minamoto
Received: 31 May 2024
Accepted: 29 November 2024
Published: 30 December 2024

Citation: Klymus KE, Baker JD, Abbott CL, Brown RJ, Craine JM, Gold Z, Hunter ME, Johnson MD, Jones DN, Jungbluth MJ, Jungbluth SP, Lor Y, Maloy AP, Merkes CM, Noble R, Patin NV, Sepulveda AJ, Spear SF, Steele JA, Takahashi M, Watts AW, Theroux S (2024) The MIEM guidelines: Minimum information for reporting of environmental metabarcoding data. Metabarcoding and Metagenomics 8: e128689. https://doi.org/10.3897/mbmg.8.128689

Introduction

Environmental DNA (eDNA) methods have become increasingly popular in the past two decades due to various advantages compared to traditional methods (Darling and Mahon 2011; Schenekar 2023). In particular, eDNA methods have been shown to have enhanced sensitivity compared to traditional methods (Lodge et al. 2012; Furlan et al. 2016; Qu and Stewart 2019; Fediajevaite et al. 2021), enabling non-invasive and non-destructive monitoring for target taxa. The high scalability of eDNA methods allow increased spatial and temporal scale monitoring, refining the resolution of distributional data (Cristescu and Hebert 2018; Garlapati et al. 2019; Veilleux et al. 2021; Morisette et al. 2021). In response to its popularity, eDNA research has become a field of its own with annual conference sessions, workshops and societies across the globe (US: Stepien et al. 2022; Stepien et al. 2023, UK: Handley et al. 2023, Australia/New Zealand: Southern eDNA Society, Japan: eDNA Society). However, the adoption of eDNA tools for real-world applications has been hindered by the lack of standardised sampling, analytical, and data reporting methods. This results in eDNA publications that contain highly inconsistent documentation of methods and results (Nicholson et al. 2020; Shea et al. 2023; Takahashi et al. 2023). Readers and peer-reviewers alike are thus faced with having to interrogate the quality and comparability of eDNA data across publications without a consistent framework for doing so.

There is momentum in the eDNA community to develop both formal standards (e.g. Gagné et al. (2021); Abbott et al. (2023)) and less formal working guidelines for species-specific and targeted eDNA approaches (e.g. Borchardt et al. (2021); Bruce et al. (2021); De Brauwer et al. (2022a)). Targeted eDNA approaches leverage quantitative or digital polymerase chain reaction (qPCR and dPCR, respectively), often for invasive and endangered species monitoring programmes (e.g. Lodge et al. (2012); Doi et al. (2017); Hunter et al. (2018); Postaire et al. (2020); Dimond et al. (2022); Nolan et al. (2023)) and have received some attention related to consistent reporting (e.g. Abbott et al. (2021)). Environmental DNA metabarcoding is used for multi-species or community level surveys and there are multiple review papers that have established a foundation of best practices for sample collection and processing, including sample volume, filter type, extraction method and primer selection (e.g. Deiner et al. (2017); Lear et al. (2018); Minamoto et al. (2021); Bruce et al. (2021); Patin and Goodwin (2023)). However, since multiple workflows can satisfy these best practices, adequate reporting is essential to ensure the data can be traced to the methods used to generate them.

There is currently a lack of guidance on metabarcoding data reporting for publication. In order for metabarcoding data to be findable, accessible, interoperable and reproducible (FAIR, Wilkinson et al. (2016)), adequate documentation of each step in the workflow as well as sample-related data and metadata are required (Shea et al. 2023). Minimum information reporting requirements published 15 years ago for qPCR data (Bustin et al. 2009) can be used as a model; here we suggest researchers and editors who publish environmental metabarcoding studies follow minimum guidelines of methods and results reporting. These guidelines cover the full range of processes in the metabarcoding workflow from field sample collection to accessibility of final datasets (Fig. 1).

The metabarcoding workflow is similar across a variety of starting sample types often used in biodiversity research, some of which are not typically referred to as eDNA. Here we follow Ruppert et al. (2019) by including samples collected either from environmental matrices (i.e. water, soil, air etc.) or from bulk, organismal sampling techniques (e.g. plankton tow nets, malaise trap etc.). We also address methods relevant to eRNA metabarcoding workflows. Thus, we use the term environmental nucleic acid (eNA) when a process is similar for both (Littlefair et al. 2022; Bunholi et al. 2023). Finally, we provide a checklist (Table 1, Suppl. material 1: file S1) and example entries (Suppl. material 1: file S2) for data and metadata reporting as a common resource for both authors and reviewers of eNA metabarcoding studies. This checklist indicates what details would be beneficial if reported in a metabarcoding study to support open science principles. We do not make recommendations on specific methodological approaches as these will vary depending on study objectives and are addressed by existing published guidelines, papers and standards. For instance, we discuss a number of controls that, if used, should be reported, but we do not make recommendations on which controls to use, as controls will vary depending on the study's objectives. The checklist provided offers a practical tool to help the community consistently and clearly report critical workflow elements, which, as an important quality measure, will expedite the adoption of eNA data for decision support.

Figure 1. Diagram of an environmental metabarcoding study workflow, including the field (blue), wet laboratory (green) and dry laboratory (orange) components. Created with BioRender.com.

Table 1. A checklist for reporting data and metadata associated with environmental metabarcoding studies. Requirements include: 'Report' - steps should be reported for FAIR (findable, accessible, interoperable and reproducible) data practices, 'If applicable' - report if component is relevant to study, 'Report for eRNA metabarcoding' - reporting of additional steps specific to eRNA studies, 'Optional' - can be reported. Reporting of these data and metadata will maximise study reproducibility and FAIR data practices. See Suppl. material 1: file S2 for a checklist that includes examples.

Step	Reporting Requirements	Reported
Methods – Laboratory Space		
General laboratory space layout	Report	
Contamination mitigation efforts in the laboratory	Report	
Methods - Metabarcoding Assay		
Target gene name	Report	
Target amplicon length	Report	
Target taxa	Report	
Primers, sequence, reference, modifications to published primers	Report	
Assay validation	Report	
Methods - Environmental Sample Collection		
Number of field samples	Report	
Definition of field sample	Report	
Number of field sample replicates	If Applicable	
Sampling dates	Report	
Sampling times	If Applicable	
Sampling locations/Geographic coordinates and geodetic datum used/Non-disclosure statement	Report	
Capture methods and materials	Report	
Sample processing method	Report	
Volume/Mass of sample collected	Report	
Volume/Mass of sample processed	Report	
Sample preservation method	Report	
Sample storage conditions	Report	
Sample storage duration	Report	
Environmental parameters	If Applicable	
Contamination mitigation efforts in the field	Report	
Positive and Negative (Site/Field/Process) controls	Report	
Methods - Nucleic Acid Extraction		
Extraction method (if using a commercial kit, provide name of kit and manufacturer)	Report	
Changes/modification to a published method or kit	Report	
Amount of sample extracted	Report	
Extraction storage conditions and duration	Report	
Nucleic acid quantification for each sample and method used to quantify	If Applicable	
Any subsequent clean-up methods	If Applicable	
Negative extraction controls	Report	
Positive extraction controls	If Applicable	
Methods - Inhibition Detection and Mitigation		
PCR inhibition detection and mitigation steps	If Applicable	
Methods - PCR Amplification and Library Preparation		
Library preparation method	Report	
Number of PCR replicates per sample	If Applicable	
Are PCR replicates pooled and indexed or indexed separately	If Applicable	
Thermal cycling instrument and manufacturer	Optional	
Thermal cycling conditions (temperatures, time at temperature, # cycles, annealing temperature(s))	Report	
Master mix composition: Final reaction volume (µl)	Report	
Master mix composition: name and manufacturer	Report	

Step	Reporting Requirements	Reported
Master mix composition: Taq concentration (X)	Report	
Master mix composition: Final concentration of each primer (forward and reverse)	Report	
Master mix composition: Volume of water added (µl)	Report	
Master mix composition: Volume (µl) or concentration of template NA	Report	
Master mix composition: Any additives (manufacturers and volumes (µl) or concentrations)	If Applicable	
Amplicon visualisation: Method used	If Applicable	
PCR clean-up methods	Report	
Size selection methods	If Applicable	
Size selection: Instrumentation and manufacturer	If Applicable	
Normalisation: Method used	If Applicable	
PCR controls (Positive PCR, Mock Community, Negative PCR or No-template) and indicate whether or not controls were also sequenced	Report	
Methods - PCR Amplification and Library Preparation (Reverse Transcription)		
Reverse transcriptase reaction kit and manufacturer	Report for eRNA metabarcoding	
Reverse transcriptase reaction conditions (primers, cycling conditions, controls, amount of template)	Report for eRNA metabarcoding	
RNA enrichment methods	Report for eRNA metabarcoding	
DNA contamination evaluation and mitigation methods	Report for eRNA metabarcoding	
Methods – Sequencing		
Sequencing instrument/ platform	Report	
Sequencing chemistry kit and manufacturer	Report	
Sequencing quality control steps	Report	
PhiX and percentage (If using Illumina platform)	If Applicable	
Methods - Bioinformatics and Reference Database		
Database creation: Source of sequences and steps to identify locus of interest	Report	
Database creation: Method for sequence curation	Report	
Database creation: Link to database or repository	Report	
Primer removal (trimming): programme, version, parameters	Report	
QC programme: programme, version, parameters	Report	
Read pair merging: programme, version, parameters	Report	
Chimera removal: programme, version, parameters	Report	
Clustering: OTUs or ASVs (and thresholds)	Report	
Additional filtering: removal of singletons or other methods	If Applicable	
Additional filtering: decontamination using sequenced controls	If Applicable	
Taxonomic assignment method	Report	
Taxonomic assignment parameters (and thresholds)	Report	
Read normalisation: methods	If Applicable	
Results - Sequencing Summary Statistics		
Total number of raw sequence reads produced	Report	
Total number of reads assigned to MIDs (i.e. tags, indices, barcodes)	Report	
Total number of reads that made it through bioinformatic filtering	Report	
Total number of reads used for final/subsequent analyses	Report	
Total number of OTUs or ASVs assigned to taxa (and to what level of taxonomy)	Report	
Total number of OTUs or ASVs unassigned	Optional	
Average number of reads per sample	Optional	
Minimum and maximum number of reads per sample	Optional	
Results from Controls (Negative/Positive/Mock Communities)	Report	
Results - Data Archiving and Availability		
Software and code archiving	If Applicable	
Raw sequence data archiving	Report	
Processed data archiving	If Applicable	

General quality control reporting

Several elements should be considered and evaluated before beginning a metabarcoding study (for in-depth review, see Bruce et al. (2021); De Brauwer et al. (2022a)). Clean laboratory practices and metabarcoding assay design are components outside of the metabarcoding workflow itself but are crucial, as well as positive and negative controls to evaluate the risk of poor assay optimisation, contamination or inhibition. Here, we discuss recommendations for reporting metadata associated with general quality control.

Laboratory spaces and practices

A key consideration is the prevention of contamination when handling eNA samples. This is particularly important in the laboratory setting where PCR is central to library preparation steps. As PCR leads to the amplification of billions of amplicons, samples, reagents, consumables and benchtops can easily become sources of contamination (Persing 1991; Aslanzadeh 2004; Willerslev and Cooper 2005). For a thorough review of clean laboratory procedures for eNA work in general, see Goldberg et al. (2016) and Patin and Goodwin (2023). It is important to describe the laboratory environment in which samples will be processed, extracted and sequenced, to enable an assessment of potential contamination risks.

The methods sections of published eNA metabarcoding studies should detail specific measures used to prevent laboratory-based contamination, such as: (1) if a unidirectional workflow was used for the wet-laboratory steps (i.e. physically separate laboratory spaces for pre- and post-PCR such that products of later steps are not introduced into spaces from earlier steps); (2) laboratory cleaning protocols and reagents (e.g. sodium hypochlorite solution); (3) workspace designated equipment and consumables; and (4) other laboratory-based contamination prevention measures including positive air pressure, HEPA-filtered air and UV-treatment of workspaces and/or consumables. We acknowledge that not all of these measures will be employed, but we point to ones that should be reported if used. This level of detail is intended to increase transparency and overall confidence in the handling of highly sensitive eNA samples.

Metabarcoding assay used

Numerous metabarcoding assays (here we define assay as a molecular analysis) have been developed targeting a range of organisms from environmental samples (Takahashi et al. 2023). When reporting on the assay used, include the target gene region, primer sequences (if newly developed) or a citation to the primer source (if previously published), any modifications made to an already published assay, expected target amplicon length and taxonomic coverage. As phylogeographic variation in target taxa may lead to primer bias or failed amplification in some species, additional validation of an assay (new or already published) is often conducted (Thalinger et al. 2021). Results of any assay validation conducted in the study should be reported. Validation of the assay should be thoroughly documented, including any *in silico*, *in vitro* and *in situ* testing, as well as results from tests demonstrating primer specificity to the targeted

group (Taberlet et al. 2018; Thalinger et al. 2021; De Brauwer et al. 2023). For further guidance on assay development, we point readers to Thalinger et al. (2021) who developed a validation scale for species-specific assays, as well as to recent guidelines for metabarcoding assays developed by the Southern eDNA Society (De Brauwer et al. 2022b, 2023).

Controls

Control samples are used throughout the metabarcoding workflow and detailed reporting of all control samples used is vital. Positive controls are generally included to ensure any run or reaction failures or anomalies (i.e. unrelated to the samples themselves) are detected (Yeh et al. 2018; Gold et al. 2022), whereas negative controls detect contamination (Borchardt et al. 2021). For eRNA studies, a negative reverse transcriptase control should be included to test for genomic DNA.

Table 2 and Fig. 2 list various controls used in a metabarcoding workflow. Not all of these will be used in every study and additional controls may be used. For instance, positive extraction controls can be used to assess extraction efficiency, but are often not used in general metabarcoding studies. Regardless, information on the number and type of controls should be provided. Controls can be sequenced to assist with interpreting final results; if they are, all relevant data and associated metadata should be reported. It is also critical to report any corrections made to final datasets using results of sequenced controls. Sepulveda et al. (2020), Bruce et al. (2021) and Takahashi et al. (2023) provide guidance on integration of controls into project design.

Metabarcoding workflow reporting

Sample collection, processing and preservation

Reporting on sample collection for eNA metabarcoding studies allows readers to evaluate the adequacy of the sampling design in the context of the study goals and limitations. A field sample should be defined by the authors. For instance, it could be a singular collection obtained during a single sampling event (i.e. a sample collected at a single location at a single point in time) or a composite sample in which material from multiple locations or times has been combined. Studies may also take replicate field samples given the inevitable stochasticity of whether the target molecule is captured by the sampling if present in the environment. Methods used for capturing samples should be reported, along with sample volume or mass, the number of field samples, field sample replicates and field control samples. In addition, if any pre-extraction processing steps are used after sample collection, the amount of sample processed and methods used (e.g. pre-filtration and/or filtration steps) should be reported. Preservation of samples, storage and duration prior to extraction should be reported as well.

Any contamination mitigation efforts made in the field, including collection of field, site and process controls should also be reported. For recommendations on sample collection and design, see Deiner et al. (2017), Dickie et al. (2018), Bowers et al. (2021), Bruce et al. (2021), Minamoto et al. (2021), De Brauwer et al. (2022a) and Pawlowski et al. (2022).

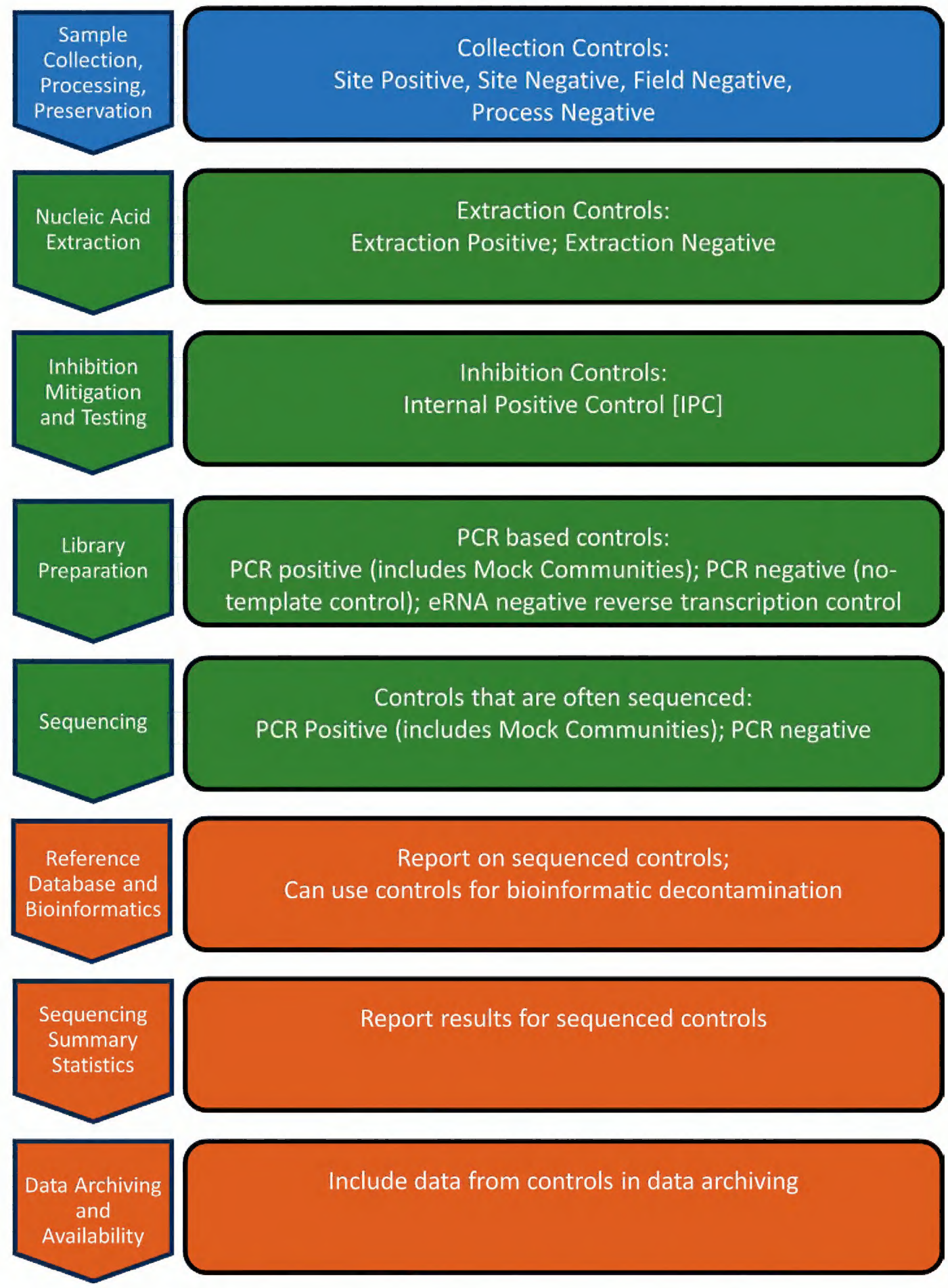
Sample Collection, Processing, Preservation
Collection Controls:
Site Positive, Site Negative, Field Negative, Process Negative
Nucleic Acid Extraction
Extraction Controls:
Extraction Positive; Extraction Negative
Inhibition Mitigation and Testing
Inhibition Controls:
Internal Positive Control [IPC]
Library Preparation
PCR based controls:
PCR positive (includes Mock Communities); PCR negative (no-template control); eRNA negative reverse transcription control
Sequencing
Controls that are often sequenced:
PCR Positive (includes Mock Communities); PCR negative
Reference Database and Bioinformatics
Report on sequenced controls;
Can use controls for bioinformatic decontamination
Sequencing Summary Statistics
Report results for sequenced controls
Data Archiving and Availability
Include data from controls in data archiving

Figure 2. Chart of the different elements of the environmental metabarcoding workflow showing the control samples involved at each step.

Detailed information describing the sampling date and locations for all samples should be included. Names of the geographical location of sampling can be reported; however, to ensure reproducibility, geographic coordinates are preferred. The latitude and longitude should be included, as well as the geodetic datum used following Geographic Information Standards (https://www.fgdc.gov/standards). Exceptions for non-disclosure of geographic coordinates are acceptable for privacy of information, cultural concerns or security reasons necessitating exact locations be withheld. Environmental parameters (e.g. temperature, salinity, pH, habitat type) should be reported, if taken, either in a table, supplemental material or within the main text. For further guidance on reporting sampling data, see standards and guidelines developed through Darwin Core (Wieczorek et al. 2012) and the Global Biodiversity Information Facility (GBIF) (Abarenkov et al. 2023).

Table 2. Table of controls that can be used throughout the metabarcoding workflow. Positive controls confirm a process is working and negative controls ensure any contamination is detected. NA refers to nucleic acid (i.e. DNA or RNA).

Control	How it's created	Why it's used
Site Positive	Collection of an environmental sample at a site where the target species is known to be present	Confirms assay viability in an environmental sample
Extraction Positive	A laboratory contrived sample to which NA or tissue is added during the NA extraction process	Introduces target NA to monitor NA extraction efficiency; used for laboratory quality control; not generally used in studies
Internal Positive Control (IPC)	A known concentration of synthetic or natural NA added to the PCR	Amplification of IPC DNA above expected cycle threshold (qPCR) suggest samples are inhibited; not generally used in metabarcoding, but see text
PCR Positive	A laboratory contrived sample to which synthetic or natural NA is added during PCR setup; can include mock community samples	Introduces target material to monitor for PCR success
Site Negative	Collection of an environmental sample at a site where the target species is known to be absent	Monitors potential non-specific amplification of the assay from the environment
Field Negative	Collection of a blank sample that follows field collection protocols	Monitors potential contamination in field collection
Process Negative	A sample added during processing that lacks target NA	Monitors potential contamination during sample processing
Extraction Negative	A laboratory contrived sample that only includes reagents used during the NA extraction process	Monitors potential contamination during extraction process
PCR Negative (No-template control; NTC)	A laboratory contrived sample that includes only PCR reagents and molecular grade water replaces the NA input	Monitors potential contamination during PCR and sequencing
Negative Reverse Transcription Control	A sample of RNA extract carried through subsequent steps, but without reverse transcriptase	Tests for DNA contamination in the RNA extract

Nucleic acid extraction

Extraction protocols are selected based on the sample, tissue or matrix targeted and what steps are required to disrupt biological membranes and release nucleic acids into solution for subsequent purification. Thus, the extraction method should be chosen, based on its efficiency for the particular sample type and/or target taxa and be reported (Djurhuus et al. 2017; Pawlowski et al. 2022). When reporting commercial kit-based extractions, the kit name and manufacturer should be included. Sample volume or mass used for extraction should be reported along with any modifications to commercial kit-based protocols or published methods. In addition, any positive and negative controls included in each extraction batch to ensure quality control of this step should be reported. Quantifying eNA post-extraction is commonly done; however, this measure will reflect total DNA concentration and not just the targeted DNA. For some substrates (e.g. water), the extraction may not yield quantifiable amounts of eNA, but will still be suitable for PCR and downstream steps. Information on how eNA extracts were stored before further use, including buffers or ethanol used and storage temperature and duration, should be reported.

Inhibition mitigation and testing

Environmental samples are prone to containing molecular compounds and trace metals that can inhibit enzymatic reactions such as PCR, thereby reducing amplification efficiency (Wilson 1997; Schrader et al. 2012; Lance and Gaun 2020; Sidstedt et al. 2020). During nucleic acid (NA) extraction, some of these inhibitory compounds may be removed, but NA dilution (McKee et al. 2015) or additional purification steps may be required for further removal (Hunter et al. 2018). An additional way to mitigate the effects of inhibitory compounds

is the use of additives to the PCR such as dimethyl sulphoxide (DMSO) and Bovine Serum Albumin (BSA) (Kreader 1996; Farell and Alexandre 2012). The use of internal positive controls (IPCs) can detect PCR inhibition in a sample prior to library preparation. This is common in qPCR studies of eDNA samples and less common in metabarcoding studies; however, there are examples of the latter (e.g. Shirazi et al. (2021) used qPCR to perform inhibition testing on all extractions in their DNA metabarcoding study). Furthermore, the U.S. Fish and Wildlife Service tests for inhibition in all metabarcoding samples collected for the Early Detection and Monitoring programme in the Great Lakes Basin (USFWS 2024). As PCR inhibition can severely reduce amplification, subsequent sequencing and data analysis can be affected; thus, any information on PCR inhibition detection and mitigation should be reported if used.

Library preparation

The metabarcoding workflow requires extracted NA samples go through a series of processing steps called library preparation before high-throughput sequencing. Libraries consist of pooled amplicon sequences that have been modified to allow for simultaneous sequencing of high numbers of samples (and subsequent demultiplexing). These modifications include addition of: sequencing primer regions; adaptors compatible to the sequencing platform; and MIDs or multiplex identifiers, that allow identification of sequences to individual samples (GBIF DNA derived data – Extension (2024)). These MIDs are often referred to as “tags”, “indices” or “barcodes” in the literature (see Deiner et al. (2017)). Outlined below we discuss the steps of library preparation (PCR, Reverse Transcription for eRNA, Post-PCR and Normalisation) and make reporting recommendations. For a more detailed explanation of steps, see Taberlet et al. (2018), Bruce et al. (2021) and Bohmann et al. (2022).

PCR

There are multiple methodologies for library preparation and the chosen method should be reported. The three predominant strategies when using Illumina instrumentation are a one-step PCR-based, two-step PCR-based or a ligation-based approach; for a detailed overview of each strategy, see Taberlet et al. (2018) and Bohmann et al. (2022). Regardless of method, the main steps in library preparation include amplification of target NA and addition of required sequencing modifications to resulting amplicons. Different PCR chemistries combined with varying thermocycling conditions may produce significantly different results due to varying amplification biases, species drop-outs or off-target amplification (Gohl et al. 2016; Gold et al. 2023; Shelton et al. 2023) and, thus, must be documented. If a kit is used during library preparation, the name and manufacturer should be reported as well as any deviations from the manufacturer’s protocol.

For all PCR steps in library preparation, chemistries and conditions should be reported. These include: total reaction volume; final concentration of primers; concentrations of master mix components and polymerase; volume of any additives; and amount of template DNA. For mastermix and polymerases, the name and manufacturer should be included. The type of thermal cycler and

manufacturer should be reported as well as thermal cycling conditions used. These conditions include the temperature and time at temperature for: the initial denaturation, denaturation, annealing, extension and final extension steps, as well as the number of cycles performed and if annealing temperature varied (e.g. touchdown PCR; Korbie and Mattick 2008).

Studies indicate multiple PCR replicates of the same sample are beneficial for environmental metabarcoding studies as they can influence species detection and richness estimates (Ficetola et al. 2015; Alberdi et al. 2018; Shirazi et al. 2021; Van den Bulcke et al. 2021). Other studies indicate that pooling multiple samples (biological replicates) may be more effective at increasing species detection and richness (Beentjes et al. 2019; Macher et al. 2021; Stauffer et al. 2021). The use of such technical or biological replicates should be described as well as whether they are indexed separately (i.e. given different MIDs) or pooled.

Controls should be included and reported in library preparation to assess potential for cross contamination of samples during all PCR steps. Using negative PCR controls allows the user to identify any laboratory contamination and positive PCR controls verify the assay is working. Mock community samples consist of known concentrations of DNA from multiple species and can provide information on amplification efficacy and bias as well as any contamination during library preparation and sequencing (Hänfling et al. 2016; Parada et al. 2016; Yeh et al. 2018; Marinchel et al. 2023).

Reverse transcription (additional step for eRNA)

For eRNA metabarcoding studies, reporting on steps from sample collection through sequencing and data reporting are the same as for eDNA; however, eRNA requires a couple of additional steps in the library preparation phase. Specifically, conversion of RNA to complementary DNA (cDNA) by reverse transcription and post-extraction enrichment of specific RNAs should be reported (Bustin and Nolan 2004; Zhao et al. 2014; Telzrow et al. 2021). For the reverse transcription reaction, the same details as described above for PCR should be reported. A concern for eRNA studies is the presence of genomic DNA contamination (see Li et al. (2022)); therefore, it is important to report detection of contaminating genomic DNA and mitigation steps taken (Laurell et al. 2012; Padhi et al. 2016; Hashemipetroudi et al. 2018; Verwilt et al. 2020).

Post-PCR

Throughout the library preparation process, sample visualisation is often done to assess the size, distribution and quantity of PCR products, as well as template clean-ups to remove unwanted components. Authors should report methods for DNA product verification as well as where in the process it was done and the type, model and manufacturer of instrumentation used.

One of the primary functions of post-PCR library preparation is to ensure amplified PCR products represent mostly the target sequences; this is done by PCR clean-up and size selection approaches. Size selection can remove non-target DNA products, primer dimer and leftover primer (Zizka et al. 2019). Size selection is important because long metabarcoding primers can be prone

to forming dimers (Peng et al. 2015) and assays may amplify non-target DNA (Collins et al. 2019). This can cause library normalisation to be biased as the final concentration will be normalised to non-target sequences, preventing detection of the targeted taxonomic group and challenges for sequencing (Alberdi et al. 2018). There are several approaches for conducting size selection, with some of the most common including the use of magnetic beads, Pippin Prep Instrumentation (Sage Science Inc., Beverly, Massachusetts, USA) and agarose gel extraction. If size selection is performed, authors should report the approach used, associated protocols, settings and parameters and manufacturers of instrumentation used. If additional PCR clean-ups are conducted, information on the methods or kits should be reported.

Library normalisation

A typical last step before sequencing a metabarcoding library is to normalise the concentration of each sample so each is represented equally in the final pooled library sequenced and will have similar read depths (Rohland and Reich 2012). Note, however, that negative control samples included in the library will have low or no quantifiable DNA and, thus, cannot be added at equal concentration (see Bruce et al. (2021) for more detail). The two main approaches used for normalisation include: 1) using DNA binding/magnetic beads to normalise and purify amplicons to a single concentration without the need to quantify; or 2) quantifying each sample and then diluting to the desired volume. Quantification can be achieved with a number of methods (see Bruce et al. (2021)). It is important to report both the normalisation method used and how sample concentration was estimated, if applicable.

Sequencing

Sequencing-by-synthesis (SBS) instruments have become the dominant instrumentation for amplicon sequencing (Bohmann et al. 2022). Platform choice depends on the number of samples, desired read depth per sample and amplicon length. Although Illumina is the most commonly used sequencing platform as of the time of writing, other technologies are appearing on the market. New platforms from Element Biosciences (San Deigo, California, USA) and Singular Genomics (San Diego, California, USA) are designed to compete with Illumina NextSeq output levels. Long-read sequencing platforms are also available and have grown in popularity (e.g. Pacific Biosciences, Menlo Park, California, USA; Oxford Nanopore Technologies, Oxford, United Kingdom).

When using short-read platforms, including Illumina and Element G4 instruments, it is recommended by the manufacturer to spike-in a PhiX sequencing control (Illumina, Inc., San Diego, California, USA; https://singulargenomics.com/g4/reagents/). For 'low diversity' libraries (such as those associated with amplicon metabarcoding), PhiX can improve sequencing quality control and provides a measure of overall run performance. If using PhiX, the percentage concentration used should be reported.

If sequencing is outsourced, information from the external facility on what quality control is performed should be obtained before releasing data. For example, most facilities will demultiplex sequence data into sample-specific files and

remove sequencing adapter sequences; they may also remove PhiX reads. As with all automated laboratory steps, the instrument platform, sequencing chemistry kit and sequencing quality control steps should be reported.

Reference database and bioinformatics

The process by which raw sequence data are converted to taxon observations and/or counts has many steps, each of which can impact results (e.g. Furlan et al. 2020; Brandt et al. 2021; De Wolfe and Wright 2023). There are currently dozens of bioinformatic tools and pipelines available (see Hakimzadeh et al. 2023 for review) and, while no pipeline is considered 'best', users should report the bioinformatic programmes, versions and steps within which to provide transparency about bioinformatic processing. We strongly encourage full reporting of bioinformatic approaches used to process and analyse eNA sequences including, but not limited to software, scripts, parameters, configuration files and variable files to allow for the repeatability of the bioinformatic analyses. Deiner et al. (2017) explain and review these various bioinformatic processes in detail, which we also outline below.

Database source and/or curation

The comprehensiveness (taxonomic breadth) and curation of reference sequence databases (Richardson et al. 2020; Gold et al. 2021; Dziedzic et al. 2023; Jeunen et al. 2023) are vital to the accurate taxonomic assignment of metabarcoding sequences (Keck et al. 2023). Thus, it is critical to include detailed information on the generation and curation of custom databases used to ensure reproducibility of the taxonomic classification (Curd et al. 2024). Such details should include: where reference sequences were acquired (GenBank, BOLD etc.); steps used to select only the locus of interest (e.g. *in silico* PCR, key word search etc.); curation methodology applied (e.g. geographic, lowest common ancestor, dereplication etc.); and a link to a data repository containing the custom reference databases used. If a custom reference database containing sequences not publicly available was used, researchers should clearly indicate this. Alternatively, many studies use large and minimally curated databases such as GenBank (Benson et al. 2013). In either case, the method, database and any database curation should be reported.

Bioinformatic processing

Multiple steps are part of the bioinformatic workflow used to convert raw sequence data into biologically analysable data (i.e. a table of sequence read numbers and associated taxonomic assignments). These steps are outlined below and include: Primer removal, Sequence quality control, Read merging, Chimera removal, OTU or ASV creation, Taxonomic assignment, Additional data filtering, and Normalisation of read data. Software workflows that encompass multiple steps exist and include MOTHUR (Schloss et al. 2009), DADA2 (Callahan et al. 2016), QIIME2 (Bolyen et al. 2019), DNAFLOW (Mousavi-Derazmahalleh et al. 2021), iMETA (Liu et al. 2023) and REVAMP (McAllister et al. 2023). If such a programme is used, the version and input parameters for each step should be reported.

Primer removal

Reads produced by sequencing platforms have fixed start positions defined by primers and include the gene region of interest. Failing to remove primers may interfere with taxonomic assignments. Some commonly used programmes for this include CUTADAPT (Martin 2011), TRIMMOMATIC (Bolger et al. 2014) and ATROPOS (Didion et al. 2017). The name and version of the programme, along with parameters and thresholds used, should be reported for all primer removal and trimming steps.

Sequence quality control

After primer removal, sequence read quality controls can be filtered initially based on minimum Q-score values, which are generated by the sequencing instrument for every nucleotide. Bioinformatic programmes like CUTADAPT and TRIMMOMATIC can be used to set quality score thresholds. The quality control programme used and associated parameters should be reported.

Read merging

When paired-end sequencing is performed, forward and reverse reads can be merged to generate a complete amplicon sequence. Off-target amplification and unremoved primer dimers can result in amplicons with different lengths than expected, which can be bioinformatically filtered out by setting a length threshold. Any reads not passing the threshold will subsequently fail to merge and be removed from the dataset; thus, read type (e.g. forward, reverse, unmerged paired, merged paired), software and parameters used (e.g. minimum number of nucleotide overlap or number of mismatches allowed) should be reported.

Chimera removal

Chimeras form during PCR and occur when two different parent DNA strands anneal together to create PCR artefacts that do not exist in nature. Chimeras can be difficult to identify (Ashelford et al. 2005) and may result in inaccurate estimates of diversity. Not all chimera products can be removed by size selection; however, many bioinformatic programmes and pipelines can identify chimeras using either *de novo* or referenced-based methods. The programme used for chimera removal and cut-off scores should be reported.

OTU or ASV creation

Operational taxonomic units (OTUs) or amplicon sequence variants (ASVs) are created from sequencing reads, which reduces the overall size of datasets and computational power needed to analyse them (Schloss and Westcott 2011; Callahan et al. 2016). OTUs are created by clustering sequences with similar identity (using a threshold chosen by the user, often at 95–99%) (Westcott and Schloss 2015). The creation of ASVs uses denoising methods (rather than dissimilarity thresholds) that detect and remove sequencing errors and distinguish sequence variants differing by as little as one nucleotide (Callahan et al.

2017). The choice of ASVs vs. OTUs should be reported as it can have a significant impact on alpha and beta diversity metrics (Chiarello et al. 2022). For an in-depth review of OTU clustering methods, see Westcott and Schloss (2015).

Taxonomic assignment

There are many methods to assign taxonomy to OTUs/ASVs. The most well-known and longest supported method is the sequence similarity-based top BLAST hit approach (Altschul et al. 1997). This uses both global and local alignments to directly compare query sequences to sequences in a reference database. Another method, sequence composition-based, assigns taxonomy based on k-mer frequencies in the query and reference database sequences. The most widely used sequence composition classifier is the Ribosomal Database Project (RDP; Wang et al. (2007)) or Naïve Bayes-based classifier. The RDP classifier can be trained for any marker and reference sets already exist for the COI animal barcode marker as well as the prokaryote 16S, fungal ITS and LSU rDNA regions (Wang et al. 2007; Cole et al. 2009; Liu et al. 2012; Porter et al. 2014; Porter and Hajibabaei 2018). Other methods of taxonomy assignment include phylogeny-based methods (Pipes and Nielsen 2024) and probabilistic-based methods such as multinomial regression (Somervuo et al. 2016). See Porter and Hajibabaei (2018) and Hleap et al. (2021) for a list of software programmes and methods commonly used.

Taxonomic assignment of some reads may not be resolvable to species level for a number of reasons. When percentage identity parameters fall below a user-defined threshold or if the reference database is sparsely populated, a single OTU/ASV might be assigned to multiple species. Furthermore, higher taxonomic levels such as genus or family may be targeted for assignment rather than at the species level. When reporting on taxonomy assignment, include methods used, threshold parameters, the reference database(s) and the taxonomic ranks assigned to each OTU/ASV. For more information on taxonomic assignment, we refer readers to Hleap et al. (2021) and Keck et al. (2023) for best practices and how to handle problems that can arise in this step.

Additional data filtering

Bioinformatic decontamination, not to be confused with laboratory decontamination, conducts post-processing quality control of OTU/ASV tables. No consensus exists on how to filter data using sequenced PCR controls. This process may employ a range of steps, including cut-offs based on sample sequencing read depth, OTU/ASV prevalence across samples, OTU/ASV proportional abundance and reads identified in the positive and negative PCR controls (De Barba et al. 2014; Hänfling et al. 2016; Kelly et al. 2018; Gold et al. 2022). Any methods, thresholds and justification for removal of reads, OTU/ASVs or samples should be reported.

Normalisation of read data

Normalisation of read data is the process by which reads are scaled or transformed to allow more accurate comparisons across samples. Normalising high throughput sequencing read data can help account for biases that occur during

PCR and sequencing (Weiss et al. 2017). Given the compositional nature of metabarcoding, the number of reads per sample may be highly variable within a sequencing run and not representative of true biological variation (Gloor et al. 2017; Willis 2019; Schloss 2024). There are many ways of normalising read data (e.g. relative abundance, eDNA index, rarefaction, Hellinger transformation), such that, authors may choose to employ different normalisation steps across different analyses. Therefore, specific normalisation methods and programme versions should be reported.

Sequencing summary statistics

Initial summary metrics from a sequencing run are often unreported, but are helpful to evaluate sequencing efficiency and evaluate metabarcoding results. They also indicate whether library preparation methods were highly effective or if changes might be necessary for future studies. Relevant summary metrics to report are: total number of raw reads; number or percentage of reads assigned to MIDs; total number or percentage of reads that passed bioinformatic filter thresholds (i.e. quality control, trimming and merging); number or percentage of reads assigned to taxonomy; the taxonomic level of assignments (i.e. species level, family level, phyla level); total number or percentage of reads unassigned to taxonomy; and total number of reads used in the final analysis or any subsequent analyses. Additionally, reporting per-sample metrics such as average number of reads per sample and minimum and maximum number of reads per sample allows an evaluation of both sequencing depth and evenness across samples.

The practice of sequencing control samples (both negative and positive) is highly recommended in the literature as deviations from expected results may indicate issues that should be addressed or accounted for either computationally or in data interpretation (Bell et al. 2017; Zinger et al. 2019; van der Loos and Nijland 2020). Reporting results from sequencing of negative controls provides transparency about potential contamination in sample processing. Likewise, read numbers from positive controls or mock communities can validate laboratory and bioinformatic procedures, as well as indicate possible false positive detections in field samples (Hänfling et al. 2016; Hleap et al. 2021). As discussed in the General Quality Control Reporting section, we acknowledge that the types and numbers of control samples taken depend on a study's goals and constraints; regardless, reporting on the use of controls, sequencing results from controls and how these reads were accounted for in the bioinformatic decontamination step provide valuable quality control information on a study.

Data archiving and availability

Alongside peer-review publications, data sharing is a core pillar of the sciences. Raw sequence data should be deposited in the International Nucleotide Sequence Database Collaboration (INSDC) nucleotide sequence archives (e.g. NCBI, ENA, DDBJ). Data and metadata should adhere to established standards such as the Minimum information about a marker gene sequence (MIMARKS) and minimum information about any sequence specifications (MixS) developed by Genomic Standards Consortium (Yilmaz et al. 2011). The use of MIxS checklists significantly improves the availability of methods and metadata and

consistency of vocabulary in archived datasets, allowing interoperability and reusability for future studies (Hassenrück et al. 2021). Processed data can also be archived using similar data repository structures (e.g. Dryad, Zenodo, Figshare) to ensure all analytical steps, especially taxonomic assignment, can be reproduced (Deiner et al. 2017). Such efforts are becoming standard in scientific publication with the proliferation of open-source code sharing platforms (e.g. GitLab, SourceForge, GitHub etc.).

Although significant bottlenecks to achieving completely FAIR metabarcoding data practices remain, at minimum, all data (i.e. OTU/ASV table, sequencing data, metadata), methods and code used to generate, analyse and interpret metabarcoding data should be provided to open-access repositories to support open science principles and enhance trust and reproducibility. Metabarcoding studies generate valuable biodiversity data and FAIR practices will allow biodiversity monitoring at increased speeds and scales, which is needed given current global biodiversity loss. All efforts should be made to adhere to the growing consensus to serve biodiversity data in large international biodiversity repositories (e.g. Global Biodiversity Information Facility (GBIF Data Standards) and Ocean Biodiversity Information System (OBIS)). In particular, we point users to Djurhuus et al. (2017), Abarenkov et al. (2023) and Silliman et al. (2023), whose articles provide all data needed to fully reproduce the study and ensure resultant biodiversity datasets follow FAIR principles. For an overview and examples of where and how to archive different datasets and metadata, see the NOAA Omics Study Data Management Guide (2024). For downloadable eDNA survey data templates, see NOAA Omics Study Data Templates (2024) that incorporate MIMARKS criteria (Yilmaz et al. 2011) and Better Biomolecular Ocean Practices protocol templates (BeBOP-OBON 2024), based on Minimum Information about an Omics Protocol (MIOP, Samuel et al. 2021).

Discussion

The advent of DNA metabarcoding has transformed our ability to census and assess biological communities. With this new capacity for generating biological data at increasing sensitivity and scale comes a deluge in environmental DNA research datasets, hence it is important that we pause and take stock of what minimum metadata should accompany environmental metabarcoding publications. Here, we identified a suite of sampling, analytical and data archiving information that should be included in publications to meet FAIR data standards and provide context for eNA results to be repeatable and interpretable. We recommend authors report these in the manuscript, supplemental materials or online resources linked to the publication (e.g. GitHub, protocols.io. etc.). This is crucial for the use and reuse of eNA data in global scale biomonitoring efforts (Berry et al. 2021; Chavez et al. 2021). Furthermore, as eNA metabarcoding methods become more routinely adopted by experts and non-experts alike, users must be able to adequately evaluate and communicate methods and data.

We recognise that the generation and curation of metabarcoding data is time and labour-intensive and that analyses require substantial computational resources and bioinformatic expertise. This can severely limit the ability of the metabarcoding community to process data quickly and efficiently into actionable biodiversity information (Shea et al. 2023), which only adds emphasis to

the need addressed in this study to maximise the usefulness of all metabarcoding datasets generated by ensuring complete and transparent reporting. The MIEM guidelines provided here on minimum information for reporting of environmental metabarcoding data parallel several similar publications by the Genomics Standards Consortium on minimum information requirements for various types of genomic data (e.g. Samuel et al. 2021; Bowers et al. 2017). Future work could support the development of additional resources to ensure truly FAIR metabarcoding data, including: 1) programmatic tools to facilitate ease of data management; 2) international metabarcoding standards via the appropriate International Standards Organisation committee (i.e. ISO/TC 147/SC 5/WG 13); and 3) consensus on best practices for data, methods and software archiving, linking and sharing. As the field continues to develop and rapidly advance, these proposed minimum reporting guidelines may be refined or updated with additional parameters. This enhanced reporting will allow for improved assessment of eNA studies during peer-review and interpretation of information for natural resource decisions.

Acknowledgements

The authors thank their respective research groups, collaborators and the eDNA community whose collective experiences led to the idea for this manuscript. We also thank Dr. Freya Rowland for assistance with figures and Dr. Brittany Perrotta for helpful comments on this paper. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service.This publication represents NOAA Pacific Marine Environmental Laboratory Contribution No. 5641.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

No funding was reported.

Author contributions

Conceptualization: Katy Klymus, Zachary Gold, Susanna Theroux. Visualization: Jacoby Baker, Katy Klymus. Writing- original draft: Katy Klymus, Jacoby Baker, Cathryn Abbott, Rachel Brown, Joseph Craine, Zachary Gold, Margaret Hunter, Mark Johnson, Devin Jones, Michelle Jungbluth, Sean Jungbluth, Yer Lor, Aaron Maloy, Christopher Merkes, Rachel Noble, Nastassia Patin, Adam J Sepulveda, Stephen Spear, Joshua Steele, Miwa Takahashi, Alison Watts, Susanna Theroux. Writing- review & editing: Susanna Theroux, Cathryn Abbott, Katy Klymus. Supervision/ Admisitration: Katy Klymus, Susanna Theroux.

Author ORCIDs

Katy E. Klymus https://orcid.org/0000-0002-8843-6241
Jacoby D. Baker https://orcid.org/0000-0002-0673-7535
Cathryn L. Abbott https://orcid.org/0000-0002-5314-7351
Rachel J. Brown https://orcid.org/0000-0001-5353-715X
Joseph M. Craine https://orcid.org/0000-0001-6561-3244
Zachary Gold https://orcid.org/0000-0003-0490-7630
Margaret E. Hunter https://orcid.org/0000-0002-4760-9302
Mark D. Johnson https://orcid.org/0000-0002-0460-945X
Devin N. Jones https://orcid.org/0000-0001-9215-2930
Michelle J. Jungbluth https://orcid.org/0000-0001-9339-7497
Sean P. Jungbluth https://orcid.org/0000-0001-9265-8341
Yer Lor https://orcid.org/0000-0002-5738-2412
Aaron P. Maloy https://orcid.org/0000-0003-4412-2280
Christopher M. Merkes https://orcid.org/0000-0001-8191-627X
Rachel Noble https://orcid.org/0000-0001-9071-8312
Nastassia V. Patin https://orcid.org/0000-0001-8522-7682
Adam J. Sepulveda https://orcid.org/0000-0001-7621-7028
Stephen F. Spear https://orcid.org/0000-0001-8351-9382
Joshua A. Steele https://orcid.org/0000-0001-8023-8956
Miwa Takahashi https://orcid.org/0000-0001-8952-051X
Alison W. Watts https://orcid.org/0000-0001-9700-6393
Susanna Theroux https://orcid.org/0000-0002-9812-7856

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Abarenkov K, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson RN, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C, Frøslev TG (2023) Publishing DNA-derived data through biodiversity data platforms, v1.3. Copenhagen: GBIF Secretariat. https://doi.org/10.35035/doc-vf1a-nr22

Abbott C, Coulson M, Gagné N, Lacoursière-Roussel A, Parent GJ, Bajno R, Dietrich C, May-McNally S (2021) Guidance on the use of targeted environmental DNA (eDNA) analysis for the management of aquatic invasive species and species at risk. DFO Canadian Science Advisory Secretariat, Ottawa, Ontario, Canada, iv + 42 pp.

Abbott C, Bright D, Bryant H, Côté G, Crookes S, Gurney K, Hanner R, Helbing C, Hocking M, Khan I, Langlois VS, Lemay M, Marshall N, Miliano R, Mirabzadeh-Ardakani A, Parent G, Richter C, Wagener A, Wilson C, Clogg-Wright K (2023) Performance criteria for the analyses of environmental DNA by targeted quantitative polymerase chain reaction. National standard of Canada, CSA W219:23. Canadian Standards Association, Toronto, Ontario, Canada, 29 pp. https://www.csagroup.org/store/product/CSA_W219%3A23/

Alberdi A, Aizpurua O, Thomas M, Gilbert P, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9(1): 134–147. https://doi.org/10.1111/2041-210X.12849

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search pro-

grams. Nucleic Acids Research 25(17): 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Applied and Environmental Microbiology 71(12): 7724–7736. https://doi.org/10.1128/AEM.71.12.7724-7736.2005

Aslanzadeh J (2004) Preventing PCR amplification carryover contamination in a clinical laboratory. Annals of Clinical and Laboratory Science 34(4): 389–396.

BeBOP-OBON (2024) BeBOP-OBON. https://github.com/BeBOP-OBON [01 May 2024]

Beentjes KK, Speksnijder AGCL, Schilthuizen M, Hoogeveen M, van der Hoorn BB (2019) The effects of spatial and temporal replicate sampling on eDNA metabarcoding. PeerJ 7: e7335. https://doi.org/10.7717/peerj.7335

Bell KL, Fowler J, Burgess KS, Dobbs EK, Gruenewald D, Lawley B, Morozumi C, Brosi BJ (2017) Applying pollen DNA metabarcoding to the study of plant-pollinator interactions. Applications in Plant Sciences 5(6): apps.1600124. https://doi.org/10.3732/apps.1600124

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Research 41(D1): D36–D42. https://doi.org/10.1093/nar/gks1195

Berry O, Jarman S, Bissett A, Hope M, Paeper C, Bessey C, Schwartz MK, Hale J, Bunce M (2021) Making environmental DNA (eDNA) biodiversity records globally accessible. Environmental DNA 3(4): 699–705. https://doi.org/10.1002/edn3.173

Bohmann K, Elbrecht V, Carøe C, Bista I, Leese F, Bunce M, Yu DW, Seymour M, Dumbrell AJ, Creer S (2022) Strategies for sample labeling and library preparation in DNA metabarcoding studies. Molecular Ecology Resources 22(4): 1231–1246. https://doi.org/10.1111/1755-0998.13512

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37(8): 852–857. https://doi.org/10.1038/s41587-019-0209-9 [Erratum in: Nature Biotechnology 37(9): 1091.]

Borchardt MA, Boehm AB, Salit M, Spencer SK, Wigginton KR, Noble RT (2021) The Environmental Microbiology Minimum Information (EMMI) Guidelines: qPCR and dPCR quality and reporting for environmental microbiology. Environmental Science & Technology 55(15): 10210–10223. https://doi.org/10.1021/acs.est.1c01767

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D et al. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35(8): 725–731. https://doi.org/10.1038/nbt.3893

Bowers HA, Pochon X, von Ammon U, Gemmell N, Stanton J-AL, Jeunen G-J, Sherman CDH, Zaiko A (2021) Towards the optimization of eDNA/eRNA sampling technologies for marine biosecurity surveillance. Water (Basel) 13(8): 1113. https://doi.org/10.3390/w13081113

Brandt MI, Trouche B, Quintric L, Günther B, Wincker P, Poulain J, Arnaud-Haond S (2021) Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. Molecular Ecology Resources 21(6): 1904–1921. https://doi.org/10.1111/1755-0998.13398

Bruce K, Blackman R, Bourlat SJ, Hellström AM, Bakker J, Bista I, Bohmann K, Bouchez A, Brys R, Clark K, Elbrecht V, Fazi S, Fonseca V, Hänfling B, Leese F, Mächler E, Mahon AR, Meissner K, Panksep K, Pawlowski J, Schmidt Yáñez P, Seymour M, Thalinger B, Valentini A, Woodcock P, Traugott M, Vasselon V, Deiner K (2021) A practical guide to DNA-based methods for biodiversity assessment. Pensoft Advanced Books, Sofia, Bulgaria, 90 pp. https://doi.org/10.3897/ab.e68634

Bunholi IV, Foster NR, Casey JM (2023) Environmental DNA and RNA in aquatic community ecology: Toward methodological standardization. Environmental DNA 5(6): 1133–1147. https://doi.org/10.1002/edn3.476

Bustin SA, Nolan T (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. Journal of Biomolecular Techniques 15(3): 155–166.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4): 611–622. https://doi.org/10.1373/clinchem.2008.112797

Callahan B, McMurdie P, Rosen M, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Callahan B, McMurdie P, Holmes S (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11(12): 2639–2643. https://doi.org/10.1038/ismej.2017.119

Chavez FP, Min M, Pitz K, Truelove N, Baker J, LaScala-Grunewald D, Blum M, Walz K, Nye C, Djurhuus A, Miller RJ, Goodwin KD, Muller-Karger FE, Ruhl HA, Scholin CA (2021) Observing life in the sea using environmental DNA. Oceanography (Washington, D.C.) 34(2): 102–119. https://doi.org/10.5670/oceanog.2021.218

Chiarello M, McCauley M, Villéger S, Jackson CR (2022) Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. PLoS ONE 17(2): e0264443. https://doi.org/10.1371/journal.pone.0264443

Cole, JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The ribosomal database project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37(Database): D141–D145. https://doi.org/10.1093/nar/gkn879

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution 10(11): 1985–2001. https://doi.org/10.1111/2041-210X.13276

Cristescu ME, Hebert PDN (2018) Uses and misuses of environmental DNA in biodiversity science and conservation. Annual Review of Ecology, Evolution, and Systematics 49(1): 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

Curd EE, Gal L, Gallego R, Silliman K, Nielsen S, Gold Z (2024) rCRUX: A rapid and versatile tool for generating metabarcoding reference libraries in R. Environmental DNA 6(1): e489. https://doi.org/10.1002/edn3.489

Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. Environmental Research 111(7): 978–988. https://doi.org/10.1016/j.envres.2011.02.001

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. Molecular Ecology Resources 14(2): 306–323. https://doi.org/10.1111/1755-0998.12188

De Brauwer M, Chariton A, Clarke LJ, Cooper MK, DiBattista J, Furlan E, Giblot-Ducray D, Gleeson D, Harford A, Herbert S, MacDonald AJ, Miller A, Montgomery K, Mooney T, Noble LM, Rourke M, Sherman CDH, Stat M, Suter L, West KM, White N, Villacorta-Rath C, Zaiko A, Trujillo-Gonzalez A (2022a) Environmental DNA protocol development guide for biomonitoring. National eDNA Reference Centre, Canberra, Australia, 49 pp. https://doi.org/10.25607/OBP-1853

De Brauwer M, Chariton A, Clarke LJ, Cooper MK, DiBattista J, Furlan E, Giblot-Ducray D, Gleeson D, Harford A, Herbert S, MacDonald AJ, Miller A, Montgomery K, Mooney T, Noble LM, Rourke M, Sherman CDH, Stat M, Suter L, West KM, White N, Villacorta-Rath C, Zaiko A, Trujillo-Gonzalez A (2022b) Environmental DNA test validation guidelines. National eDNA Reference Centre, Canberra, Australia, 33 pp. https://doi.org/10.25607/OBP-1886

De Brauwer M, Clarke LJ, Chariton A, Cooper MK, de Bruyn M, Furlan E, MacDonald AJ, Rourke ML, Sherman CDH, Suter L, Villacorta-Rath C, Zaiko A, Trujillo-González A (2023) Best practice guidelines for environmental DNA biomonitoring in Australia and New Zealand. Environmental DNA 5(3): 417–423. https://doi.org/10.1002/edn3.395

De Wolfe TJ, Wright ES (2023) Multi-factorial examination of amplicon sequencing workflows from sample preparation to bioinformatic analysis. BMC Microbiology 23(1): 107. https://doi.org/10.1186/s12866-023-02851-8

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–5895. https://doi.org/10.1111/mec.14350

Dickie IA, Boyer S, Buckley HL, Duncan RP, Gardner PP, Hogg ID, Holdaway RJ, Lear G, Makiola A, Morales SE, Powell JR, Weaver L (2018) Towards robust and repeatable sampling methods in eDNA-based studies. Molecular Ecology Resources 18(5): 940–952. https://doi.org/10.1111/1755-0998.12907

Didion JP, Martin M, Collins FS (2017) Atropos: Specific, sensitive, and speedy trimming of sequencing reads. PeerJ 5: e3720. https://doi.org/10.7717/peerj.3720

Dimond JL, Gathright BR, Bouma JV, Carson HS, Sowul K (2022) Detecting endangered pinto abalone (*Haliotis kamtschatkana*) using environmental DNA: Comparison of ddPCR, qPCR, and conventional diver surveys. Environmental DNA 4(6): 1397–1406. https://doi.org/10.1002/edn3.351

Djurhuus A, Port J, Closek CJ, Yamahara KM, Romero-Maraccini O, Walz KR, Goldsmith DB, Michisaki R, Breitbart M, Boehm AB, Chavez FP (2017) Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. Frontiers in Marine Science 4: 314. https://doi.org/10.3389/fmars.2017.00314

Doi H, Katano I, Sakata Y, Souma R, Kosuge T, Nagano M, Ikeda K, Yano K, Tojo K (2017) Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. Royal Society Open Science 4(7): 170568. https://doi.org/10.1098/rsos.170568

Dziedzic E, Sidlauskas B, Cronn R, Anthony J, Cornwell T, Friesen TA, Konstantinidis P, Penaluna BE, Stein S, Levi T (2023) Creating, curating and evaluating a mitogenomic reference database to improve regional species identification using environmental DNA. Molecular Ecology Resources 23(8): 1880–1904. https://doi.org/10.1111/1755-0998.13855

Farell EM, Alexandre G (2012) Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. BMC Research Notes 5(1): 257. https://doi.org/10.1186/1756-0500-5-257

Fediajevaite J, Priestley V, Arnold R, Savolainen V (2021) Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. Ecology and Evolution 11(9): 4803–4815. https://doi.org/10.1002/ece3.7382

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources 15(3): 543–556. https://doi.org/10.1111/1755-0998.12338

Furlan EM, Gleeson D, Hardy CM, Duncan RP (2016) A framework for estimating the sensitivity of eDNA surveys. Molecular Ecology Resources 16(3): 641–654. https://doi.org/10.1111/1755-0998.12483

Furlan EM, Davis J, Duncan RP (2020) Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid zone waterholes. Molecular Ecology Resources 20(5): 1259–1276. https://doi.org/10.1111/1755-0998.13170

Gagné N, Bernatchez L, Bright D, Côté G, Coulson M, Gurney K, Hanner R, Helbing C, Hobbs J, Hocking M, Khan I, Naumann C, Parent G, Richter C, Silverio C, Skinner M, Weir A, Wilcox T, Wilson C, Clogg-Wright K (2021) Environmental DNA (eDNA) reporting requirements and terminology. National standard of Canada, CSA W214:21. Canadian Standards Association, Toronto, Ontario, Canada, 31 pp. https://www.csagroup.org/store/product/CSA%20W214%3A21/

Garlapati D, Charankumar B, Ramu K, Madeswaran P, Ramana Murthy MV (2019) A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Reviews in Environmental Science and Biotechnology 18(3): 389–411. https://doi.org/10.1007/s11157-019-09501-4

GBIF DNA derived data – Extension (2024) GBIF DNA derived data – Extension. https://rs.gbif.org/extension/gbif/1.0/dna_derived_data_2022-02-23.xml [19 March 2024]

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome datasets are compositional: And this is not optional. Frontiers in Microbiology 8: 2224. https://doi.org/10.3389/fmicb.2017.02224

Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, Knights D, Beckman KB (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nature Biotechnology 34(9): 942–949. https://doi.org/10.1038/nbt.3601

Gold Z, Curd EE, Goodwin KD, Choi ES, Frable BW, Thompson AR, Walker Jr HJ, Burton RS, Kacev D, Martz LD, Barber PH (2021) Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. Molecular Ecology Resources 21(7): 2546–2564. https://doi.org/10.1111/1755-0998.13450

Gold Z, Wall AR, Schweizer TM, Pentcheff ND, Curd EE, Barber PH, Meyer RS, Wayne R, Stolzenbach K, Prickett K, Luedy J, Wetzer R (2022) A manager's guide to using eDNA

metabarcoding in marine ecosystems. PeerJ 10: e14071. https://doi.org/10.7717/peerj.14071

Gold Z, Shelton AO, Casendino HR, Duprey J, Gallego R, Van Cise A, Fisher M, Jensen AJ, D'Agnese E, Andruszkiewicz Allen E (2023) Signal and noise in metabarcoding data. PLoS ONE 18(5): e0285674. https://doi.org/10.1371/journal.pone.0285674

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS, Laramie MB, Mahon AR, Lance RF, Pilliod DS, Strickler KM, Waits LP, Fremier AK, Takahara T, Herder JE, Taberlet P (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution 7(11): 1299–1307. https://doi.org/10.1111/2041-210X.12595

Hakimzadeh A, Abdala Asbun A, Albanese D, Bernard M, Buchner D et al. (2023) A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. Molecular Ecology Resources 00: 1–17. https://doi.org/10.1111/1755-0998.13847

Handley LL, Blackwell T, Broadhurst HA, Clark K, Davison PI, England J, Mariani S, McDevitt AD, Pillay K, Read D, Walsh K, Nisbet A, Creer S (2023) UK DNA working group eDNA week, January 2022. Environmental DNA 5(1): 18–24. https://doi.org/10.1002/edn3.364

Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25(13): 3101–3119. https://doi.org/10.1111/mec.13660

Hashemipetroudi SH, Nematzadeh G, Ahmadian G, Yamchi A, Kuhlmann M (2018) Assessment of DNA contamination in RNA samples based on ribosomal DNA. Journal of Visualized Experiments 131(131): e55451. https://doi.org/10.3791/55451-v

Hassenrück C, Poprick T, Helfer V, Molari M, Meyer R, Kostadinov I (2021) FAIR enough? A perspective on the status of nucleotide sequence data and metadata on public archives. bioRxiv, 1–24. https://doi.org/10.1101/2021.09.23.461561

Hleap JS, Littlefair JE, Steinke D, Hebert PDN, Cristescu ME (2021) Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. Molecular Ecology Resources 21(7): 2190–2203. https://doi.org/10.1111/1755-0998.13407

Hunter M, Meigs-Friend G, Ferrante J, Kamla AT, Dorazio R, Keith-Diagne L, Luna F, Lanyon J, Reid J (2018) Surveys of environmental DNA (eDNA): A new approach to estimate occurrence in vulnerable manatee populations. Endangered Species Research 35: 101–111. https://doi.org/10.3354/esr00880

Jeunen G-J, Lamare M, Cummings V, Treece J, Ferreira S, Massuger J, Pryor Rodgers L, Tait L, Lust B, Wilkinson S, Mariani S, Mills S, Gemmell N (2023) Unveiling the hidden diversity of marine eukaryotes in the Ross Sea: A comparative analysis of seawater and sponge eDNA surveys. Environmental DNA 5(6): 1780–1792. https://doi.org/10.1002/edn3.500

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Kelly RP, Gallego R, Jacobs-Palmer E (2018) The effect of tides on nearshore environmental DNA. PeerJ 6: e4521. https://doi.org/10.7717/peerj.4521

Korbie D, Mattick J (2008) Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature Protocols 3(9): 1452–1456. https://doi.org/10.1038/nprot.2008.133

Kreader CA (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Applied and Environmental Microbiology 62(3): 1102–1106. https://doi.org/10.1128/aem.62.3.1102-1106.1996

Lance RF, Guan X (2020) Variation in inhibitor effects on qPCR assays and implications for eDNA surveys. Canadian Journal of Fisheries and Aquatic Sciences 77(1): 23–33. https://doi.org/10.1139/cjfas-2018-0263

Laurell H, Iacovoni JS, Abot A, Svec D, Maoret JJ, Arnal JF, Kubista M (2012) Correction of RT-qPCR data for genomic DNA-derived signals with ValidPrime. Nucleic Acids Research 40(7): e51. https://doi.org/10.1093/nar/gkr1259

Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckly TR, Cruickshank R, Dopheide A, Handley KM, Hermans S, Kamke J, Lee CK, MacDiarmid R, Morales SE, Orlovich DA, Smissen R, Wood J, Holdaway R (2018) Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. New Zealand Journal of Ecology 42(1): 10. https://doi.org/10.20417/nzjecol.42.9

Li X, Zhang P, Wang H, Yu Y (2022) Genes expressed at low levels raise false discovery rates in RNA samples contaminated with genomic DNA. BMC Genomics 23(1): 554. https://doi.org/10.1186/s12864-022-08785-1

Littlefair JE, Rennie MD, Cristescu ME (2022) Environmental nucleic acids: A field-based comparison for monitoring freshwater habitats using eDNA and eRNA. Molecular Ecology Resources 22(8): 2928–2940. https://doi.org/10.1111/1755-0998.13671

Liu KL, Porras-Alfaro A, Kuske CR, Eichorst SA, Xie G (2012) Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. Applied and Environmental Microbiology 78(5): 1523–1533. https://doi.org/10.1128/AEM.06826-11

Liu Y-X, Chen L, Ma T, Li X, Zheng M, Zhou X, Chen L, Qian X, Xi J, Lu H, Cao H, Ma X, Bian B, Zhang P, Wu J, Gan R-Y, Jia B, Sun L, Ju Z, Gao Y, Wen T, Chen T (2023) EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research. iMeta 2(1): e83. https://doi.org/10.1002/imt2.83

Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, Feder JL, Mahon AR, Pfrender ME (2012) Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. Molecular Ecology 21(11): 2555–2558. https://doi.org/10.1111/j.1365-294X.2012.05600.x

Macher T-H, Schütz R, Arle J, Beermann AJ, Koschorreck J, Leese F (2021) Beyond fish eDNA metabarcoding: Field replicates disproportionately improve the detection of stream associated vertebrate species. Metabarcoding and Metagenomics 5: e66557. https://doi.org/10.3897/mbmg.5.66557

Marinchel N, Marchesini A, Nardi D, Girardi M, Casabianca S, Vernesi C, Penna A (2023) Mock community experiments can inform on the reliability of eDNA metabarcoding data: A case study on marine phytoplankton. Scientific Reports 13(1): 20164. https://doi.org/10.1038/s41598-023-47462-5

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17(1): 3. https://doi.org/10.14806/ej.17.1.200

McAllister SM, Paight C, Norton EL, Galaska MP (2023) SPOTLIGHT ON REVAMP Rapid Exploration and Visualization through an Automated Metabarcoding Pipeline. Oceanography (Washington, D.C.) 36(2/3): 114–119. https://www.jstor.org/stable/27257889

McKee AM, Spear SF, Pierson TW (2015) The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biological Conservation 183: 70–76. https://doi.org/10.1016/j.biocon.2014.11.031

Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, Araki H, Iwasaki W, Kasai A, Masuda R, Uchii K (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. Environmental DNA 3(1): 8–13. https://doi.org/10.1002/edn3.121

Morisette J, Burgiel S, Brantley K, Daniel WM, Darling J, Davis J, Franklin T, Gaddis K, Hunter M, Lance R, Leskey T, Passamaneck Y, Piaggio A, Rector B, Sepulveda A, Smith M, Stepien CA, Wilcox T (2021) Strategic considerations for invasive species managers in the utilization of environmental DNA (eDNA): Steps for incorporating this powerful surveillance tool. Management of Biological Invasions : International Journal of Applied Research on Biological Invasions 12(3): 747–775. https://doi.org/10.3391/mbi.2021.12.3.15

Mousavi-Derazmahalleh M, Stott A, Lines R, Peverley G, Nester G, Simpson T, Zawierta M, De La Pierre M, Bunce M, Christophersen CT (2021) eDNAFlow, an automated, reproducible and scalable workflow for analysis of environmental DNA sequences exploiting Nextflow and Singularity. Molecular Ecology Resources 21(5): 1697–1704. https://doi.org/10.1111/1755-0998.13356

Nicholson A, McIsaac D, MacDonald C, Gec P, Mason BE, Rein W, Wrobel J, de Boer M, Milián-García Y, Hanner RH (2020) An analysis of metadata reporting in freshwater environmental DNA research calls for the development of best practice guidelines. Environmental DNA 2(3): 343–349. https://doi.org/10.1002/edn3.81

Nolan KP, Loeza-Quintana T, Little HA, McLeod J, Ranger B, Borque DA, Hanner RH (2023) Detection of brook trout in spatiotemporally separate locations using validated eDNA technology. Journal of Environmental Studies and Sciences 13(1): 66–82. https://doi.org/10.1007/s13412-022-00800-x

NOAA Omics Data Management Guide (2024) NOAA Omics Data Management Guide. https://noaa-omics-dmg.readthedocs.io/en/latest/ [01 May 2024]

NOAA Omics Study Data Templates (2024) NOAA Omics Study Data Templates. https://noaa-omics-templates.readthedocs.io/en/latest/ [01 May 2024]

Padhi BK, Singh M, Huang N, Pelletier G (2016) A PCR-based approach to assess genomic DNA contamination in RNA: Application to rat RNA samples. Analytical Biochemistry 494: 49–51. https://doi.org/10.1016/j.ab.2015.10.012

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18(5): 1403–1414. https://doi.org/10.1111/1462-2920.13023

Patin NV, Goodwin KD (2023) Capturing marine microbiomes and environmental DNA: A field sampling guide. Frontiers in Microbiology 13: 1026596. https://doi.org/10.3389/fmicb.2022.1026596

Pawlowski J, Bruce K, Panksep K, Aguirre FI, Amalfitano S, Apothéloz-Perret-Gentil L, Baussant T, Bouchez A, Carugati L, Cermakova K, Cordier T, Corinaldesi C, Costa FO, Danovaro R, Dell'Anno A, Duarte S, Eisendle U, Ferrari BJD, Frontalini F, Frühe L, Haegerbaeumer A, Kisand V, Krolicka A, Lanzén A, Leese F, Lejzerowicz F, Lyautey E, Maček I, Sagova-Marečková M, Pearman JK, Pochon X, Stoeck T, Vivien R, Weigand A, Fazi S (2022) Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. The Science of the Total Environment 818: 151783. https://doi.org/10.1016/j.scitotenv.2021.151783

Peng Q, Satya RV, Lewis M, Randad P, Wang Y (2015) Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC Genomics 16(1): 589. https://doi.org/10.1186/s12864-015-1806-8

Persing DH (1991) Polymerase Chain Reaction: Trenches to Benches. Journal of Clinical Microbiology 29(7): 1281–1285. https://doi.org/10.1128/jcm.29.7.1281-1285.1991

Pipes L, Nielsen R (2024) A rapid phylogeny-based method for accurate community profiling of large-scale metabarcoding datasets. eLife 13: e85794. https://doi.org/10.7554/eLife.85794

Porter TM, Hajibabaei M (2018) Automated high throughput animal DNA metabarcode classification. Scientific Reports 8(1): 4226. https://doi.org/10.1038/s41598-018-22505-4

Porter TM, Gibson JF, Shokralla S, Baird DJ, Golding GB, Hajibabaei M (2014) Rapid and accurate taxonomic classification of insect (class Insecta) cytochrome c oxidase subunit 1 (COI) DNA barcode sequences using a naïve Bayesian classifier. Molecular Ecology Resources 14(5): 929–942. https://doi.org/10.1111/1755-0998.12240

Postaire BD, Bakker J, Gardiner J, Wiley TR, Chapman DD (2020) Environmental DNA detection tracks established seasonal occurrence of blacktip sharks (*Carcharhinus limbatus*) in a semi-enclosed subtropical bay. Scientific Reports 10(1): 11847. https://doi.org/10.1038/s41598-020-68843-0

Qu C, Stewart KA (2019) Evaluating monitoring options for conservation: Comparing traditional and environmental DNA tools for a critically endangered mammal. Naturwissenschaften 106(3–4): 9. https://doi.org/10.1007/s00114-019-1605-1

Richardson RT, Sponsler DB, McMinn-Sauder H, Johnson RM (2020) MetaCurator: A hidden Markov model-based toolkit for extracting and curating sequences from taxonomically-informative genetic markers. Methods in Ecology and Evolution 11(1): 181–186. https://doi.org/10.1111/2041-210X.13314

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22(5): 939–946. https://doi.org/10.1101/gr.128124.111

Ruppert K, Kline R, Rahman M (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Samuel RM, Meyer R, Buttigieg PL, Davies N, Jeffery NW, Meyer C, Pavloudi C, Pitz KJ, Sweetlove KM, Theroux S, van de Kamp J, Watts A (2021) Toward a global public repository of community protocols to encourage best practices in biomolecular ocean observing and research. Frontiers in Marine Science 8: 758694. https://doi.org/10.3389/fmars.2021.758694

Schenekar T (2023) The current state of eDNA research in freshwater ecosystems: Are we shifting from the developmental phase to standard application in biomonitoring? Hydrobiologia 850(6): 1263–1282. https://doi.org/10.1007/s10750-022-04891-z

Schloss PD (2024) Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. MSphere 9(2): e00354–e23. https://doi.org/10.1128/msphere.00354-23

Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Applied and Environmental Microbiology 77(10): 3219–3226. https://doi.org/10.1128/AEM.02810-10

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75(23): 7537–7537. https://doi.org/10.1128/AEM.01541-09

Schrader C, Schielke A, Ellerbroek L, Johne R (2012) PCR inhibitors - occurrence, properties and removal. Journal of Applied Microbiology 113(5): 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Sepulveda AJ, Hutchins PR, Forstchen M, Mckeefry MN, Swigris AM (2020) The elephant in the lab (and field): Contamination in aquatic environmental DNA studies. Frontiers in Ecology and Evolution 8: 609973. https://doi.org/10.3389/fevo.2020.609973

Shea MM, Kuppermann J, Rogers MP, Smith DS, Edwards P, Boehm AB (2023) Systematic review of marine environmental DNA metabarcoding studies: Toward best practices for data usability and accessibility. PeerJ 11: e14993. https://doi.org/10.7717/peerj.14993

Shelton AO, Gold ZJ, Jensen AJ et al. (2023) Toward quantitative metabarcoding. Ecology 104(2): e3906. https://doi.org/10.1002/ecy.3906

Shirazi S, Meyer RS, Shapiro B (2021) Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecology and Evolution 11(22): 15766–15779. https://doi.org/10.1002/ece3.8239

Sidstedt M, Rådström P, Hedman J (2020) PCR inhibition in qPCR, dPCR and MPS—Mechanisms and solutions. Analytical and Bioanalytical Chemistry 412(9): 2009–2023. https://doi.org/10.1007/s00216-020-02490-2

Silliman K, Anderson S, Storo R, Thompson L (2023) A case study in sharing marine eDNA metabarcoding data to OBIS. Biodiversity Information Science and Standards 7: e111048. https://doi.org/10.3897/biss.7.111048

Somervuo P, Koskela S, Pennanen J, Henrik Nilsson R, Ovaskainen O (2016) Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics (Oxford, England) 32(19): 2920–2927. https://doi.org/10.1093/bioinformatics/btw346

Stauffer S, Jucker M, Keggin T, Marques V, Andrello M, Bessudo S, Cheutin M-C, Borrero-Pérez GH, Richards E, Dejean T, Hocdé R, Juhel J-B, Ladino F, Letessier TB, Loiseau N, Maire E, Mouillot D, Mutis Martinezguerra M, Manel S, Fernández AP, Valentini A, Velez L, Albouy C, Pellissier L, Waldock C (2021) How many replicates to accurately estimate fish biodiversity using environmental DNA on coral reefs? Ecology and Evolution 11(21): 14630–14643. https://doi.org/10.1002/ece3.8150

Stepien CA, Theroux S, Weisberg SB (2022) The Second National Workshop on Marine eDNA: A workshop to accelerate the incorporation of eDNA science into environmental management applications. Environmental DNA 6(1): e379. https://doi.org/10.1002/edn3.379

Stepien CA, Lance RF, Klymus KE, Hunter ME (2023) The government eDNA working group 6th annual eDNA technical exchange workshop. Environmental DNA 5(6): 1196–1201. https://doi.org/10.1002/edn3.466

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For biodiversity research and monitoring. Oxford University Press, Oxford, USA, 272 pp. https://doi.org/10.1093/oso/9780198767220.001.0001

Takahashi M, Saccò M, Kestel JH, Neste G, Campbell MA, van der Heyde M, Heydenrych MJ, Juszkiewicz DJ, Nevill P, Dawkins KL, Bessey C, Fernandes K, Miller H, Power M, Mousavi-Derazmahalleh M, Newton JP, White NE, Richards ZT, Allentoft ME (2023) Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. The Science of the Total Environment 873: 162322. https://doi.org/10.1016/j.scitotenv.2023.162322

Telzrow CL, Zwack PJ, Righi SE, Dietrich FS, Chan C, Owzar K, Alspaugh JA, Granek JA (2021) Comparative analysis of RNA enrichment methods for preparation of *Cryptococcus neoformans* RNA sequencing libraries. G3 (Bethesda, Md.) 11(11): jkab301. https://doi.org/10.1093/g3journal/jkab301

Thalinger B, Deiner K, Harper LR, Rees HC, Blackman RC, Sint D, Traugott M, Goldberg CS, Bruce K (2021) A validation scale to determine the readiness of environmental DNA assays for routine species monitoring. Environmental DNA 3(4): 823–836. https://doi.org/10.1002/edn3.189

U.S. Fish and Wildlife Service (2024) Protocol for environmental DNA monitoring for fish mitochondrial 12S rRNA. U.S. Fish and Wildlife Service, Midwest Region, Bloomington, Minnesota, and Northeast Region, Hadley, MA, USA. [available upon request]

Van den Bulcke L, De Backer A, Ampe B, Maes S, Wittoeck J, Waegeman W, Hostens K, Derycke S (2021) Towards harmonization of DNA metabarcoding for monitoring marine macrobenthos: The effect of technical replicates and pooled DNA extractions on species detection. Metabarcoding and Metagenomics 5: e71107. https://doi.org/10.3897/mbmg.5.71107

Van der Loos LM, Nijland R (2021) Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology 30(13): 3270–3288. https://doi.org/10.1111/mec.15592

Veilleux HD, Misutka MD, Glover CN (2021) Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. The Science of the Total Environment 782: 146891. https://doi.org/10.1016/j.scitotenv.2021.146891

Verwilt J, Trypsteen W, Van Paemel R, De Preter K, Giraldez MD, Mestdagh P, Vandesompele J (2020) When DNA gets in the way: A cautionary note for DNA contamination in extracellular RNA-seq studies. Proceedings of the National Academy of Sciences of the United States of America 117(32): 18934–18936. https://doi.org/10.1073/pnas.2001675117

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73(16): 5261–5267. https://doi.org/10.1128/AEM.00062-07

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, Hyde ER, Knight R (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5(1): 27. https://doi.org/10.1186/s40168-017-0237-y

Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3: e1487. https://doi.org/10.7717/peerj.1487

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3(1): 160018. https://doi.org/10.1038/sdata.2016.18

Willerslev E, Cooper A (2005) Ancient DNA. Proceedings. Biological Sciences 272(1558): 3–16. https://doi.org/10.1098/rspb.2004.2813

Willis AD (2019) Rarefaction, alpha diversity, and statistics. Frontiers in Microbiology 10: 2407. https://doi.org/10.3389/fmicb.2019.02407

Wilson IG (1997) Inhibition and facilitation of nucleic acid amplification. Applied and Environmental Microbiology 63(10): 3741–3751. https://doi.org/10.1128/aem.63.10.3741-3751.1997

Yeh Y, Needham DM, Sieradzki ET, Fuhrman JA (2018) Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequencing run. mSystems 3: https://doi.org/10.1128/mSystems.00023-18

Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PS, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spo A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29(5): 415–420. https://doi.org/10.1038/nbt.1823

Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics 15(1): 419. https://doi.org/10.1186/1471-2164-15-419

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kauserud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P (2019) DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology 28(8): 1857–1862. https://doi.org/10.1111/mec.15060

Zizka VMA, Leese F, Peinert B, Geiger MF (2019) DNA metabarcoding from sample fixative as a quick and voucher preserving biodiversity assessment method. Genome 62(3): 122–136. https://doi.org/10.1139/gen-2018-0048

Supplementary material 1

Additional information

Authors: Katy E. Klymus, Jacoby D. Baker, Cathryn L. Abbott, Rachel J. Brown, Joseph M. Craine, Zachary Gold, Margaret E. Hunter, Mark D. Johnson, Devin N. Jones, Michelle J. Jungbluth, Sean P. Jungbluth, Yer Lor, Aaron P. Maloy, Christopher M. Merkes, Rachel Noble, Nastassia V. Patin, Adam J. Sepulveda, Stephen F. Spear, Joshua A. Steele, Miwa Takahashi, Alison W. Watts, Susanna Theroux

Data type: xlsx

Explanation note: **file S1.** A printable checklist of data and metadata that should be reported or reported if available for environmental metabarcoding studies. **file S2.** The checklist of data and metadata that should be reported or reported if available for environmental metabarcoding studies, including example entries for reference.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128689.suppl1